DOCUMENT RESUME

ED 075 487 TM 002 566

AUTHOR Stake, Robert E.
TITLE Responsive Evaluation.
PUB DATE 5 Dec 72
NOTE 4p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Data Collection; *Evaluation Criteria; *Evaluation Techniques; *Information Needs; *Program Evaluation; Technical Reports

ABSTRACT

The definition, structures, utilities, stimulus-response differences, and portrayals of responsive evaluation are presented. An educational evaluation is said to be a "responsive evaluation" if it orients more directly to program activities than to program intents, if it responds to audience requirements for information, and if the different value-perspectives present are referred to in reporting the success of the program. The structure of the responsive evaluation follows the primary components of the instructional program. The choice of tests and other data-gathering devices is made based on observation of the program in action and interaction with various interested groups. Responsive evaluation is particularly useful during both formative and summative evaluation. The principal stimuli of responsive evaluation are those of the program, including responses of students and subsequent dialogue. The responsive approach tries to respond to the natural ways in which people assimilate information and arrive at understanding. Program portrayal may feature descriptions of persons, often by case studies. The challenge to the evaluator is to minimize the sampling error and to find ways to authenticate this less "scientific" way of reporting. (DB)

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

ED 075487

TM 002 566

Robert E. Stake
December 5, 1972

RESPONSIVE EVALUATION

Most contemporary plans for the evaluation of educational programs are "preordinate." They rely on prespecification. They emphasize (1) statement of goals, (2) use of objective tests, (3) standards held by program personnel, and (4) research-type reports. It is even presumed by some that these are essential features of any evaluation plan. They are not. There is an important alternative to preordinate evaluation: responsive evaluation.

This is not a new alternative. Responsive evaluation is what people do naturally in evaluating things. They observe and react. What is new is the beginning of a technology developed around this natural behavior, in part to overcome its defects.

An educational evaluation is a "responsive evaluation" if it orients more directly to program activities than to program intents, if it responds to audience requirements for information, and if the different value-perspectives present are referred to in reporting the success of the program. In these three separate ways an evaluation plan can be responsive.

An evaluator is employed by a client to do an evaluation, with certain audiences in mind. To do a responsive evaluation, the evaluator conceives of a plan of observations and negotiations. He arranges for various persons to observe the program. With their help he prepares brief narratives, portrayals, product displays, graphs, etc. He finds out what of that is of value to his audience. He gathers expressions of worth from various points of view. Of course, he checks on the quality of his records. He gets program personnel to react to the accuracy of his portrayals. He gets authority figures to react to the importance of various findings. He gets audience members to react to the relevance of his findings. He does much of this informally, iterating--keeping a record of actions and reactions. He chooses media accessible to him and his audiences to increase the likelihood and fidelity of communication. He might prepare a final written report, he might not--depending on what he and his client have agreed on.

Structures

Responsive evaluations require planning and structure; but they rely little on formal statements and abstract representations, e.g., flow charts, test scores. Statements of objectives, hypotheses, test batteries, teaching syllabi are, of course, given primary attention if they are primary components of the instructional program. Then they are treated not as the basis for the evaluation plan but as components of the instructional plan. These components are to be evaluated just as other components are.

Tests and other data-gathering devices are not ruled out. The choice of these instruments is made as a result of observing the program in action and of interacting with various groups having an interest in the program.

The proper amount of structure for responsive evaluation depends on the program and persons involved. There is a need to plan and interact naturalistically. Too much and too little planning are both bad. Too much and too little prose representation are both bad. The structure should serve the purposes of the evaluation--in preordinate evaluation, structure sometimes dictates purpose.

Utilities

Responsive evaluation will be particularly useful during formative evaluation when the project staff needs help in monitoring the program and when no one is sure what problems will arise. It will be particularly useful in summative evaluation when audiences want an understanding of the activities and of the strengths and shortcomings of the program. The responsive evaluator may see his responsibility as indirectly providing a "shared experience," one that the audience cannot directly share for one reason or another.

Preordinate evaluation should be preferred to responsive evaluation when it is important to know if certain goals have been reached, if certain promises have been kept, and when certain hypotheses or issues are to be investigated. With the greater focus and opportunity for preparation, measurements made can be expected to be more objective and reliable. To the extent that aims or issues change, the preordinate approach may be less desirable.

There are many reasons why preordinate evaluation can be ineffective. It is likely to be underfunded, understaffed, and initiated too late. But even under optimum conditions it often will fail. A collection of specific objectives will understate educational purposes. Different people have different purposes. Side effects--good ones and bad--get ignored. Program background, conditions, transactions are likely to be poorly described. Standardized tests seldom match objectives, criterion referenced tests oversimplify and fail to measure transfer, and custom-built tests are poorly validated. And people cannot read many of the reports or do not find them useful.

Responsive evaluation is not likely to overcome all of these obstacles. But it is an approach that is attentive to them. There are problems with the responsive approach too. Not enough time or resources may be available to measure key outcomes. The results may be seen as too subjective. The assets and liabilities of the two approaches need to be weighed before and during an evaluation study.

S - R Differences

Probably because of his ties with the experimental psychologist, the preordinate evaluator has dealt himself a substantial responsibility for providing a standardized stimulus (e.g., behavioral objective, test item). The whole evaluation project is treated as a "stimulus" to which the teachers or students will respond, and their responses will constitute an evaluation report.

Confident in his ability to comprehend and operationalize purposes, this evaluator defines his role as one of arranging a stimulus condition (or taking advantage of an existing one) which evokes a critical performance. He records the performance objectively so as to demonstrate the effects of the instructional program.

The success of his efforts depends not only on whether or not he can devise stimuli having "criterion" value, but on his ability to report them in a useful way. He often seems so taken by the elegance of some of his arrangements that he fails to explain how the responses do and do not relate to program objectives. By getting program people to limit their objectives at the outset to what he can evoke, he simplifies his work. It is always hoped (and all too seldom challenged) that there is a high correspondence (the basis for indirect measurement) between the observed response variables and the true criterion response variables.

Were he to have had greater ties with the anthropologist, the journalist, and the poet,* the contemporary evaluator might have dealt himself a more responsive assignment. The principal stimuli then would be those of the program, including responses of students and subsequent dialogue. Somehow (hopefully not imperially) the evaluator would pick and choose what to observe, what to record. He might not be wholly passive; he might find that what is needed requires intervention, stimulation. But he would arrive at that decision (with his client, with his audiences) as a result of his letting the program (its plan, its process, its product) stimulate him. The preordinate evaluator and the responsive evaluator both do some of both, of course.

Portrayals

One of the principal reasons for backing away from the preordinate approach to evaluation is to improve communication with audiences. The responsive approach tries to respond to the natural ways in which people assimilate information and arrive at understanding. It appears that direct experience is an efficient and satisfying way of creating understanding. (The understanding may be misunderstanding, but validity is not the issue at the moment.) Our best substitute for direct experience might be surrogate experience where the observer uses attending- and organizing-paradigms similar to the audience's. Such paradigms are not likely to be those of the specialist in measurement or the theoretically minded social scientist. The surrogate experience probably will be reported best in terms of persons, places, and events.

It seems that the conventional research-report paradigm is a very unnatural way of communicating. Characteristics (descriptors, traits) are identified; and

---

*The poet and these other folks want to stimulate their readers, but they are loathe to stimulate (create phenomena) just so that they have something to write about.

relationships among them are featured. Individuals are observed, found to differ, and the distribution of scores is described. Covariations of various kinds are reported and interpreted. From such a report it is very hard and often even impossible for a reader to know what the program "was like." If he is supposed to learn what the program "was like," the evaluation report should be different from the research report.

Often the portrayal will feature descriptions of persons. The evaluator will find that "case studies" of several students may more interestingly and faithfully represent the educational program than a few measurements on all the participants. The promise of gain is two-fold. The readers will comprehend and the complexity of the program is not lost. The several participants cannot be considered a satisfactory representation of the many so a sampling error occurs. The protests about the sampling error will be loud; but the size of the error may be small, and it often will be a satisfactory price to pay for the improvement in communication.

The data collected by an evaluator are sometimes represented in a matrix such as this:

	Person					
Variable	1	2	3	4	5	6
A						
B						
C						
D						
.						
.						
.						

(Teacher, class, district project are equivalent ways of categorizing the columns.) The conventional way of aggregating the data is by rows, with scores aggregated over persons. Only a few rows are selected customarily for primary attention. Comparison of groups and correlations are considered on those few variables with individual differences a source of error, depersonalized.

For portrayal, to many audiences, the data should be aggregated by columns. Only a few columns can be given primary attention. These will be the narratives of how a few students were engaged by the instructional program, how they interacted with teachers and students, how they studied, what they learned, how they felt. It is common knowledge that this approach is useful for discussing programs. The challenge to the evaluator is to minimize the sampling error and to find ways to authenticate this more casual, less "scientific" way of reporting.